Homework #12

1. Consider the least squares problem,

$$\min_{x \in \mathbb{R}^p} \|y - Ax\|_2^2, \tag{1}$$

where A is an $n \times p$ matrix and $y \in \mathbb{R}^n$. Let x^* be a solution of this problem. (We do not assume that x^* is a unique solution.)

   (a) Consider the problem,

$$\begin{aligned} &\min_{x \in \mathbb{R}^p} \|x\|^2 \\ &\text{subject to } \|y - Ax\|_2^2 = \|y - Ax^*\|^2. \end{aligned} \tag{2}$$

   Let $A = USV^T$ be the svd decomposition of A. Let s_i be the ith singular value of A, so that $S_{ii} = s_i$. Let S^+ be the $p \times n$ diagonal matrix defined by,

$$S^+_{ii} = \begin{cases} \frac{1}{s_i} & \text{if } s_i \neq 0 \\ 0 & \text{if } s_i = 0 \end{cases} \tag{3}$$

   Show that the solution x^+ of (2) satisfies $x^+ = VS^+U^Ty$. The matrix US^+V^T is called the pseudoinverse of A. (Hint: You won't need KKT here. Instead, expand x in the V basis and y in the U basis. You'll be able to determine Ax and derive an expression for $\|y - Ax\|^2$. Then you'll be able to show that x^+ satisfies the constraint and has minimal norm.)

   (b) Consider the penalized least squares problem,

$$\min_{x \in \mathbb{R}^p} \|y - Ax\|_2^2 + \lambda \|x\|^2. \tag{4}$$

   with $\lambda \geq 0$. Let x_λ be the solution to the penalized problem. We know that for $\lambda > 0$, the solution is unique. Show that $\|x_\lambda\|$ is decreasing in λ and that $\lim_{\lambda \to 0} x_\lambda = x^+$. (Hint: You know $x_\lambda = (A^TA + \lambda I)^{-1}A^Ty$. Plug in the svd for A, write y in the U basis and then show $x_\lambda \cdot v_i = x^+ \cdot v_i$.

   (c) Go back to homework 10 and redo 3(a) using x^+. You should be able to determine whether the Lagrange multiplier satisfies $\nu = 0$ based on $\|x^+\|$. (In my solutions to hw

10, I had a lemma which noted $\|x_\lambda\| \to x_0$ or $\|x_\lambda\| \to \infty$ as $\lambda \to 0$. This is true, with $x_0 = x^+$, but you can now show that the $\|x_\lambda\| \to \infty$ case doesn't happen.)

(d) Go back to homework 10 and redo 3(c), but do not assume the existence of $(A^T A)^{-1}$.

2. The file `dataset_separable.csv` provides separable data (x_i, y_i) for $i = 1, 2, \ldots, 1000$. $x_i \in \mathbb{R}^2$ and $y_i \in \{-1, 1\}$. Recall the SVM for separable data.

$$\min_{a \in \mathbb{R}^2, b \in \mathbb{R}} \|a\|^2$$
$$\text{subject to } y_i(a^T x_i + b) \geq 1.$$

(a) Derive the dual problem. (I did this in class. I want you to go through it yourself.)

(b) Use the Python package `qpsolvers` to solve the dual problem for the dual variable ν (following our class notation). See the link below for installation and implementation instructions.

`https://pypi.org/project/qpsolvers/`

(Note: On my 2024 mac, I had some problem installing some of the qp solvers used in the package. The default solver is quadprog and seems to install without difficult. To use it set `solver="quadprog"` in your call to `solve_qp`. However, quadprog requires a positive definite quadratic term. The dual involves the term of the form $\nu^T A \nu$ with A positive semidefinite. You can shift by $A + 10^{-6} I$ to make the term positive definite.)

(c) Use ν to compute a, b and identify the support vectors (i.e. the x_i for which $\nu_i \neq 0$. Plot the data colored by the y_i with the support vectors plotted a different color. On top of the data, plot the SVM hyperplane. Verify that your accuracy is 100%. (There is no need to split into training and test here.)

3. (In this problem, we will develop the theory for kernel, non-separable SVM. To keep things manageable, you don't need to fit the SVM to a dataset. We will do that next week.)

Consider a dataset (x_i, y_i) for $i = 1, 2, \ldots, n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$. Recall the SVM for non-separable data.

$$\min_{a \in \mathbb{R}^p, b \in \mathbb{R}} \|a\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$\text{subject to } y_i(a^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0.$$

We will consider a kernel version of this problem. Let $k(x, y)$ be a reproducing kernel. Define $z_i = k(x, x_i)$ and consider the new dataset (z_i, y_i) and the kernel version of the primal problem,

$$\min_{a \in H, b \in \mathbb{R}} \|a\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$\text{subject to } y_i(a^T z_i + b) \geq 1 - \xi_i, \xi_i \geq 0.$$

Note that $a \in H$ and $z_i \in H$ where H is the RKHS associated with the kernel. Properly, H could be an infinite dimensional space, but for this problem we will treat H as a finite dimensional Euclidean space. The analysis for an infinite dimensional space is essentially the same. So you can take $H = \mathbb{R}^q$.

(a) Derive the dual problem for the kernel problem. (I essentially did this in class.)

(b) For $x \in \mathbb{R}$, show that the SVM predicts $y = \text{sign}(\sum_{i=1}^{n} \nu_i y_i k(x_i, x))$. (Note a feature of the SVM: prediction requires the use of the training data.)